

**DEVELOPMENT OF PCA-BASED FAULT DETECTION SYSTEM BASED ON  
VARIOUS OF NOC MODELS FOR CONTINUOUS-BASED PROCESS**

**MOHAMAD YUSUP BIN ABD WAHAB**

**A thesis submitted in the fulfillment of the  
requirements for the award of the degree of  
Bachelor of Chemical Engineers**

**Faculty of Chemical and Natural Resources Engineering  
University Malaysia Pahang**

**JULY, 2012**

## **ABSTRACT**

Multivariate Statistical Process Control (MSPC) technique has been widely used for fault detection and diagnosis. Currently, contribution plots are used as basic tools for fault diagnosis in MSPC approaches. This plot does not exactly diagnose the fault, it just provides greater insight into possible causes and thereby narrow down the search. Hence, the cause of the faults cannot be found in a straightforward manner. Therefore, this study is conducted to introduce a new approach for detecting and diagnosing fault via correlation technique. The correlation coefficient is determined using multivariate analysis techniques, namely Principal Component Analysis (PCA). In order to overcome these problems, the objective of this research is to develop new approaches, which can improve the performance of the present conventional MSPC methods. The new approaches have been developed, the Outline Analysis Approach for examining the distribution of Principal Component Analysis (PCA) scores, the Correlation Coefficient Approach for detecting changes in the correlation structure within the variables. This research proposed PCA Outline Analysis Control Chart for fault detection. The result from the conventional method and ne approach were compared based on their accuracy and sensitivity. Based on the results of the study, the new approaches generally performed better compared to the conventional approaches, particularly the PCA Outline Analysis Control Chart.

## ABSTRAK

Kawalan Proses Statistik multivariat (MSPC) teknik telah digunakan secara meluas untuk mengesan kerosakan dan diagnosis. Pada masa ini, plot sumbangan digunakan sebagai alat asas untuk diagnosis kerosakan dalam pendekatan MSPC. Plot ini tidak tepat mendiagnosis kerosakan, ia hanya memberikan gambaran yang lebih besar ke dalam sebab-sebab yang mungkin dan dengan itu mengecilkan carian. Oleh itu, punca kesalahan yang tidak boleh ditemui dalam cara yang jelas dan mudah. Oleh itu, kajian ini dijalankan untuk memperkenalkan satu pendekatan baru untuk mengesan dan mendiagnosis kerosakan melalui teknik korelasi. Pekali korelasi ditentukan menggunakan teknik analisis multivariat, iaitu Analisis Komponen Utama (PCA). Bagi mengatasi masalah ini, objektif kajian ini adalah untuk membangunkan pendekatan baru, yang boleh meningkatkan prestasi kaedah konvensional yang hadir MSPC. Pendekatan baru telah dibangunkan, Analisis Pendekatan Ringkasan untuk memeriksa taburan Analisis komponen prinsipal (PCA) skor, Pendekatan Korelasi Pekali untuk mengesan perubahan dalam struktur korelasi dalam pembolehubah. Kajian ini mencadangkan PCA Panjang Analisis Carta Kawalan untuk mengesan kerosakan. Hasil dari kaedah konvensional dan pendekatan ne berbanding berdasarkan ketepatan dan sensitiviti mereka. Berdasarkan hasil kajian itu, pendekatan baru yang dilakukan secara umumnya lebih baik berbanding dengan pendekatan konvensional, terutamanya Carta Kawalan Analisis Rangka PCA.

## TABLE OF CONTENT

CHAPTER	TITLE	PAGE
	<b>SUPERVISOR'S DECLARATION</b>	<b>ii</b>
	<b>STUDENT'S DECLARATION</b>	<b>iii</b>
	<b>DEDICATION</b>	<b>iv</b>
	<b>ACKNOWLEDGEMENT</b>	<b>v</b>
	<b>ABSTRACT</b>	<b>vi</b>
	<b>ABSTRAK</b>	<b>vii</b>
	<b>TABLE OF CONTENT</b>	<b>viii</b>
	<b>LIST OF TABLE</b>	<b>xi</b>
	<b>LIST OF FIGURE</b>	<b>xii</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Background of Study	1
	1.1.1 Statistical process control (SPC)	1
	1.1.2 Multivariate Statistical Process Control (MSPC)	3
	1.2 Problem Statement	6
	1.3 Objective of Study	7
	1.4 Research Question	7
	1.5 Scope of study	7
	1.6 Rational and Significance	8
	1.7 Contributions	8
	1.8 Organization of This Report	8

<b>2</b>	<b>LITERATURE REVIEW</b>	<b>10</b>
2.1	Introduction	10
2.2	Extension Of Principal Component Analysis	10
2.3	Fundamentals / Theory of Process Monitoring on MSPM Using PCA Tools	13
2.4	Extension Of Principal Component Analysis	15
2.4.1	Multiway-PCA	15
2.4.2	Multi-Block PCA	16
2.4.3	Moving PCA	17
2.4.4	Dissimilarity, DISSIM	18
2.4.5	Multi-Scale PCA	19
2.4.6	Kernel of PCA	19
2.5	Extension of Multivariate Statistical Process Monitoring	21
2.6	Summary	25
<b>3</b>	<b>METHODOLOGY</b>	<b>26</b>
3.1	Introduction	26
3.2	Phase I: Off-line Modelling and Monitoring	27
3.3	Phase II: On-line Monitoring	30
3.4	Summary	30
<b>4</b>	<b>RESULTAND DISCUSSIONS</b>	<b>31</b>
4.1	Introduction	31
4.2	Case Study of CSTR System	31
4.3	Normal Operating Condition Data Collection	33
4.4	Fault data collection	36

<b>5</b>	<b>CONCLUSION AND RECOMMENDATION</b>	<b>45</b>
5.1	Conclusion	45
5.2	Recommendation	46
	<b>LIST OF REFERENCES</b>	<b>47</b>
	<b>APPENDIX</b>	<b>49</b>

**LIST OF TABLE**

<b>TABLE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
<b>1.1</b>	Definitions of the Domains used in SPC	<b>2</b>
<b>4.1</b>	List of variables in the CSTR system for monitoring	<b>31</b>
<b>4.2</b>	List of abnormal operations in CSTR	<b>35</b>

## LIST OF FIGURE

FIGURE NO.	TITLE	PAGE
1.1	Concept of SPC	1
1.2	Concept of MSPC	5
2.1	System for Two Variables Distribution	12
2.2	Graphical Representation of PCA	12
2.3	Multi-block PCA monitoring framework	16
2.4	Linear PCA and Kernel PCA	19
3.1	Two main phases namely as off-line modelling and monitoring	26
4.1	CSTR system	30
4.2	Accumulated data variance explained by different PCs	33
4.3	T2 statistic for NOC data and SPE statistic for NOC data	33
4.4	T2 statistic for NOC data and SPE statistic for NOC data	34
4.5	Fault 3 abrupt, Fault 3 incipient , Fault 4 abrupt, Fault 4 incipient	36
4.6	Results of PCA-based MSPM system for incipient fault	37
4.7	Results of PCA-based MSPM system for Abrupt fault.	38
4.8	Results of different modes of NOC with PCA-based MSPM system for incipient fault.	39
4.9	Results of different modes of NOC with PCA-based MSPM system for Abrupt fault	39
4.10	Results of PCA-based MSPM system for incipient fault	40
4.11	Results of PCA-based MSPM system for Abrupt fault.	40
4.12	Results of different modes of NOC with PCA-based MSPM system for incipient fault.	41



<b>4.13</b>	Results of different modes of NOC with PCA-based MSPM system for Abrupt fault.	42
-------------	--	----

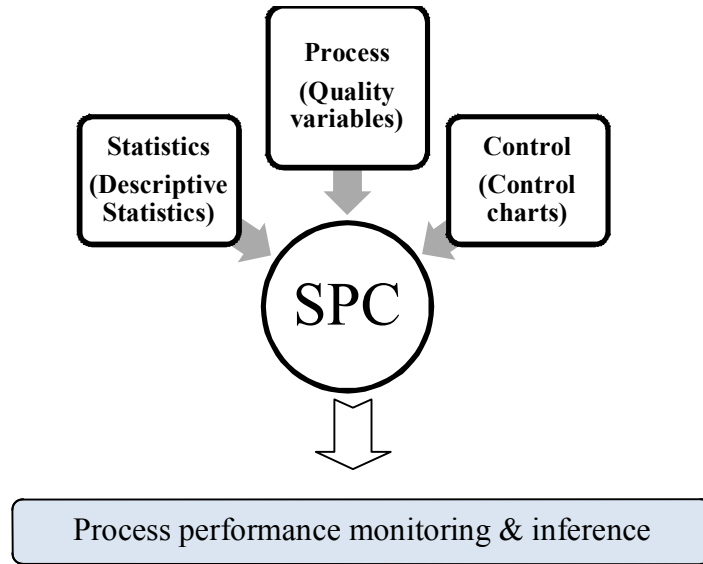
## CHAPTER 1

### INTRODUCTION

#### 1.1 Background of Study

##### 1.1.1 Statistical process control (SPC)

Statistical process control (SPC) involves using statistical techniques to measure and analyze the variation in processes. Most often used for manufacturing processes, the intent of SPC is to monitor product quality and maintain processes to fixed targets. Statistical quality control refers to using statistical techniques for measuring and improving the quality of processes and includes SPC in addition to other techniques, such as sampling plans, experimental design, variation reduction, process capability analysis, and process improvement plans. Therefore, the true meaning of SPC could also be represented conceptually by a diagram as shown in **Figure 1.1**, whereby those basic domains in **Table 1.1** have been integrated to serve as an integrated function, known as process monitoring.



**Figure 1.1:** Concept of SPC

**Table 1.1** Definitions of the Domains used in SPC

Domains	Definition
Statistical	Drawing conclusions using scientific or mathematical approach of analyzing data.
Process	The whole combination of people, equipment, materials, methods and environment working together to produce output; any work area that has identifiable, measurable output.
Control	Making something behave in a predictable consistent manner.

As the main focus is to monitor the process performance over time, functionally, it could also be used to detect those possible large process variations, or namely faults, in the quality products or outputs which can then leads to the prevention of fault sources from the inputs. In short, SPC is a useful technique especially to detect, diagnose and eventually performing process stabilization for any given process, in terms of providing an early warning for plant operators as well as gaining the deeper understanding on the process behaviour that tend to be achieved in the process. Due to this, from the

commercial and quality assurance point of view, therefore, the aims of using SPC are mainly to (Wetherill and Brown, 1991):

- i. improve quality (as a whole) by means of reduced manufacturing costs and increased customer satisfaction, and also,
- ii. to increase productivity (process and products) by maximizing output value relative to inputs.

A primary tool used for SPC is the control chart, a graphical representation of certain descriptive statistics for specific quantitative measurements of the manufacturing process. These descriptive statistics are displayed in the control chart in comparison to their "in-control" sampling distributions. The comparison detects any unusual variation in the manufacturing process, which could indicate a problem with the process. Several different descriptive statistics can be used in control charts and there are several different types of control charts that can test for different causes, such as how quickly major vs. minor shifts in process means are detected. Control charts are also used with product measurements to analyze process capability and for continuous process improvement efforts. Over time, other process-monitoring tools have been developed, including:

- Cumulative Sum (CUSUM) charts: the ordinate of each plotted point represents the algebraic sum of the previous ordinate and the most recent deviations from the target.
- Exponentially Weighted Moving Average (EWMA) charts: each chart point represents the weighted average of current and all previous subgroup values, giving more weight to recent process history and decreasing weights for older data.

### **1.1.2 Multivariate Statistical Process Control (MSPC)**

Multivariate Statistical Process Control (MSPC) (Kresta *et al.*, 1991; MacGregor, 1994; Kourti and MacGregor, 1995; MacGregor and Kourti, 1995; Kasonovich *et al.*, 1996; Wise and Gallagher, 1996; Wold *et al.*, 1999 and Lennox, 2001) is known

generally as an upgraded technique, from which, it was emerged as a result of reformation in conventional Statistical Process Control (SPC) method. The significant enhancement made to MSPC is that it takes into consideration of multivariate behaviour among all variable relations simultaneously as opposed to the conventional schemes in SPC. In order to appreciate the distinctions between MSPC and SPC fundamentally, the first section will elaborate more on this issue according to their individual basic domains respectively. Next, the essentials of MSPC implementations are then to be discussed. In addition, an overview of the advanced projects conducted for MSPC are also presented subsequently.

Therefore, in order to apply MSPC successfully, those domains of SPC should be considered as well as understood especially from the angle of multivariate standpoint. Basically, SPC concerns with one basic question and that is ‘given a sequence of data sets, did all come from the same population?’ (Wheeler and Chambers, 1990). Therefore, as to answer this question thoroughly with respect to MSPC method, there are also other several related questions need to be answered fundamentally according to those three major fields of SPC, which are:

- i. Descriptive statistics:
  - a. Where the values are centred based on multivariate manner?
  - b. How to measure the spread out of those multivariate values?
- ii. Quality Variables
  - a. How can the outcomes of the certain procedures be described when the system internally affect to each others?
  - b. What are the techniques used to represent the values described by the multivariate descriptive statistics?
- iii. Control charts
  - a. How can the properties of the sample be used to infer the stability of the given population?
  - b. What are the shapes of distribution of the values described by the multivariate descriptive statistics if the system has been randomized?
  - c. What are the criteria to distinguish between two different populations under a single sample observation?

Hence, the following discussions are focusing on to answer those questions according to the MSPC groundwork. Before continue to those sections, as the mathematical basis of any multivariate analysis is on data matrices (Green and Carroll,1976; Chatfield and Collins,1980; Marvin,1982; Cox and Cox,1994; Borg and Groenen,1997) the followings show some of the references in presenting those mathematical symbols, that they might be relevance in this report:

- i. For any given matrices, technically, they are denoted by boldface capital symbols.
- ii. Boldface lowercase letters represent form for vectors.
- iii. Every element of either matrices or vectors and other mathematical terms including scalar values will be written in the form of *italic lowercase*.

Thus, it obviously shows that SPC is not really a suitable monitoring method for any kind of multivariate process. As a result, the objectives of having MSPC is therefore should be understood based on the following arguments:

- i. MSPC is the only method, of which, the data is treated simultaneously into a single monitoring by way of reducing the dimensionality of the data observed without losing any of important information.
- ii. The technique is capable of capturing on the directionality of process variation information especially on how all the variables are behaving relative to one another. (MacGregor and Kourti, 1995).
- iii. By using the multivariate method, the presence of noise levels could be reduced through averaging (MacGregor and Kourti, 1995).
- iv. This approach can reduce the burden of constructing a large amount of single-variable control charts and enable detecting events that are impossible or difficult to detect from single-variable control charts (Phatak, 1999).

On top of those objectives, the original goals of SPC are also been considered as well as carried together, such a way that the ultimate aim is to increase productivity of high quality products for multivariate process, which is illustrated in **Figure 1.2**.

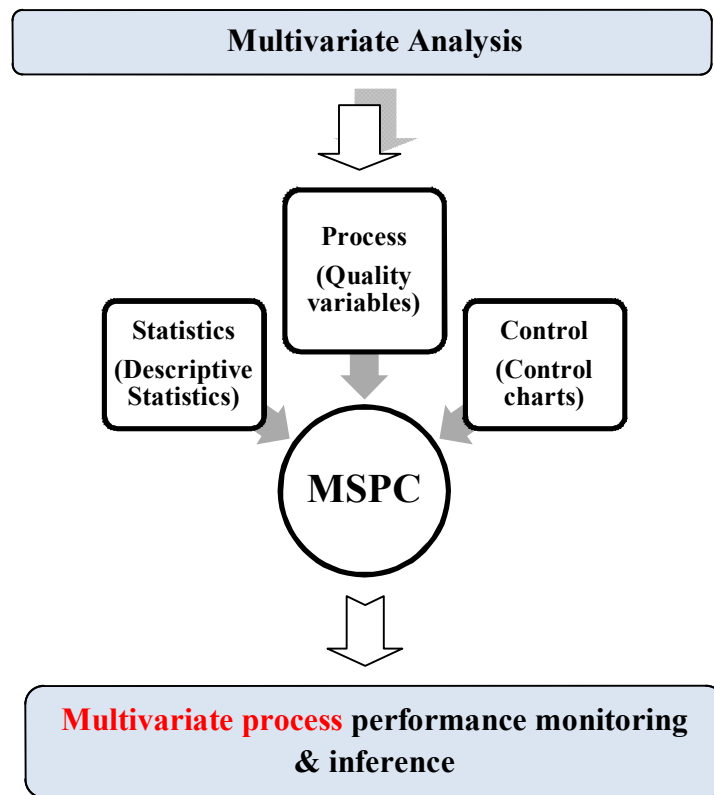


Figure 1.2: Concept of MSPC

## 1.2 Problem Statement

In order to ensure the successfulness of any operation, it is important to detect process upsets, equipment malfunctions or other special events as early as possible and then to diagnose and remove the factors that cause those events. However, Zhao et al., (2004) mentioned that a process which is having multiple operating modes tends trigger continuous warning signal even when the process itself is operating under another steady-state. SPC is rather complicated, in the sense that there are many individual control charts need to be monitored because the collection number of variables observed is also exist in a large quantity. MSPC is the only method, of which, the data is treated simultaneously into a single monitoring by way of reducing the dimensionality of the data observed without losing any of important information.

### 1.3 Objective of Study

The main purpose of this research is to study the impact of applying various modes of normal operating condition (NOC) in terms of the number of samples and variable variations on the process monitoring performance for continuous-based process. Therefore, the main objectives of this research are:

- i. To develop the conventional MSPM method based on a single NOC
- ii. To implement the conventional MSPM method based on different modes of NOC.
- iii. To analyze the monitoring performance between system (i) and (ii).

### 1.4 Research Question

- i. What is the main impact of reducing the number of samples as well as variations on the monitoring performance?
- ii. What are the criteria should be used in selecting the NOC model?

### 1.5 Scope of study

Scope of propose study are on the development of PCA-based fault detection system based on various modes of NOC models for continuous-based process. There are three main scope will be investigated using MATLAB.

- i. The conventional MSPM method will be develop based on single NOC mode. The linear PCA algorithm is used for reducing the multivariate data dimensions.
- ii. The MSPM will be run traditionally by implementing different mode, which in this research is on two modes. According to Zhao et al. (2004),in spite of the success of applying PCA based MSPM tools to process data for detecting abnormal situations, when these tools are applied to a process with multiple



operating modes, many missing and false alarms appear even when the process itself under other steady-state nominal operating conditions.

- iii. As all data have been obtained, it will be analyze further with two multivariate control charts namely Hotelling's  $T^2$  and Squared Prediction Errors (SPE) statistic for the fault detection operation.

## 1.6 Rational and Significance

In this research, effort mainly concentrates on breaking through the current limitation and the further application of MSPC on a multivariate continuous chemical process. The main contributions of this research are:

1. Application of MSPC tools on the fault detection and diagnosis.
2. An Eigenvalue-eigenvector PCA approach had been used for developing Principals Components model.
  - Modified Process Fault Detection and Diagnosis, mechanisms are also developed based on the Outline Analysis.

## 1.7 Contributions

- i. A new set of criteria is proposed for selecting the optimized NOC data for monitoring.
- ii. As a result of (i), the monitoring performance can be enhanced in terms of missing and false alarm.

## 1.8 Organization of This Report

The new monitoring algorithm has been proposed in this study by developing PCA-based fault detection system based on various modes of NOC models for continuous-based process. Hence, this report is divided into five main chapters. The first chapter discusses the background of the works which includes the problem statement, objectives, scopes and contributions. **Chapter II** which is literature review describes the

fundamental of MSPC and justification of applying PCA in MSPM frameworks. **Chapter III** explains the research methodology of this study. **Chapter IV** presents some of the preliminary results. Conclusions and further research works are given in **Chapter V**.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Introduction

According to Venkatasubramaniam, Rengaswamy, Kavuri and Yin (2003) MSPM tools are data driven technique that generally reduce the dimension of process data and extract key features and trends that are of interest to plant personnel. MSPM tools used to reduces dimensions of process data, like PCA and subsequent refinements, which have show great success. In chapter 2, we will discuss on the fundamental or theory of process monitoring on MSPM using PCA tools, process monitoring issues and extension and justification of applying PCA in MSPM frameworks. Lastly, a summary is given at the end of this chapter.

#### 2.2 Fundamental of PCA (Principal Component Analysis)

Principal Component Analysis (PCA) is one of the most common multivariate analyses applied in the MSPC area (Jackson, 1991, MacGregor and Kourti, 1995; Zhang *et al.*, 1997, Gnanadesikan, 1997). In particular, PCA can be described by means of either using mathematical representation or graphical representation. Firstly, from the mathematical point of view, PCA is a multivariate projection technique, which can transform a set of original variables  $[\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_m]$  to a set of new variables  $[\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_m]$ . Generally, these newly formed variables are called Principal Components, PC for short (Jackson, 1991), whereby they build the individual linear combinations of the original variables which is simplified as follows:

$$\mathbf{P}_{n \times m} = \mathbf{X}_{n \times m} \mathbf{V}_{m \times m} \quad (2.1)$$

Initially, the original matrix,  $\mathbf{X}$  has  $m$  variables with each variable has  $n$  number of measurements. The data are arranged in the form of  $n \times m$ , where the measurements of a variable are organized in the form of a column vector, which is shown as follows:

$$\mathbf{X}_{n \times m} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_m] = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,m} \\ x_{2,1} & x_{2,2} & \dots & x_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,m} \end{bmatrix} \quad (2.2)$$

Next,  $\mathbf{V}$  is known as the eigenvector matrix, whereby it gives the weighting function in forming the linear combinations of the original variables. The eigenvector matrix,  $\mathbf{V}$ , contains eigenvectors or also known as loading vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ . Each eigenvector  $\mathbf{v}_m$  is a column vector which contains the arrangement of elements  $\mathbf{v}_m^T = [v_{1,m}, v_{2,m}, \dots, v_{m,m}]^T$ , as denoted in the following matrix:

$$\mathbf{V}_{m \times m} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_m] = \begin{bmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,m} \\ v_{2,1} & v_{2,2} & \dots & v_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m,1} & v_{m,2} & \dots & v_{m,m} \end{bmatrix} \quad (2.3)$$

Lastly,  $\mathbf{P}$  is the principal components scores matrix, in which it contains  $n$  scores for each of the principal components, as given subsequently as:

$$\mathbf{P} = [\mathbf{p}_1 \quad \dots \quad \mathbf{p}_m] \quad (2.4)$$

$$= \begin{bmatrix} x_{1,1}v_{1,1} + x_{1,2}v_{2,1} + \dots + x_{1,m}v_{m,1} & \dots & x_{1,1}v_{1,m} + x_{1,2}v_{2,m} + \dots + x_{1,m}v_{m,m} \\ x_{2,1}v_{1,1} + x_{2,2}v_{2,1} + \dots + x_{2,m}v_{m,1} & \dots & x_{2,1}v_{1,m} + x_{2,2}v_{2,m} + \dots + x_{2,m}v_{m,m} \\ \vdots & \vdots & \vdots \\ x_{n,1}v_{1,1} + x_{n,2}v_{2,1} + \dots + x_{n,m}v_{m,1} & \dots & x_{n,1}v_{1,m} + x_{n,2}v_{2,m} + \dots + x_{n,m}v_{m,m} \end{bmatrix} \quad (2.5)$$

Thus, as far as multivariate calculations are concerned,  $\mathbf{P}$ , will play the role of quality variables rather than individual variables of  $\mathbf{X}$ , in the MSPC method. The

original data,  $\mathbf{X}$ , are also could be predictable backed from the calculated PC element, by which, the procedures are:

$$\mathbf{X}_{n \times m} \mathbf{V}_{m \times m} = \mathbf{P}_{n \times m} \quad (2.6)$$

$$\mathbf{X}_{n \times m} \mathbf{V}_{m \times m} \mathbf{V}_{m \times m}^T = \mathbf{P}_{n \times m} \mathbf{V}_{m \times m}^T \quad (2.7)$$

$$\mathbf{X}_{n \times m} \mathbf{I}_{m \times m} = \mathbf{P}_{n \times m} \mathbf{V}_{m \times m}^T \quad (2.8)$$

$$\mathbf{X}_{n \times m} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \dots \quad \mathbf{p}_m] \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_m \end{bmatrix} \quad (2.9)$$

If all the PCs are used to represent the original variables, the original raw data matrix is reproduced back as shown in equation (2.24). However, in this situation, the purpose for using Principal Components Analysis as data dimension reduction technique will be lost. In order to maintain the uniqueness of this technique, only several PCs will be used to represent most of the original data variation. Therefore, if ‘ $a$ ’ of Principal Components are decided to be retained with  $a < m$ , then equation (2.24) can be adjusted and written as follows:

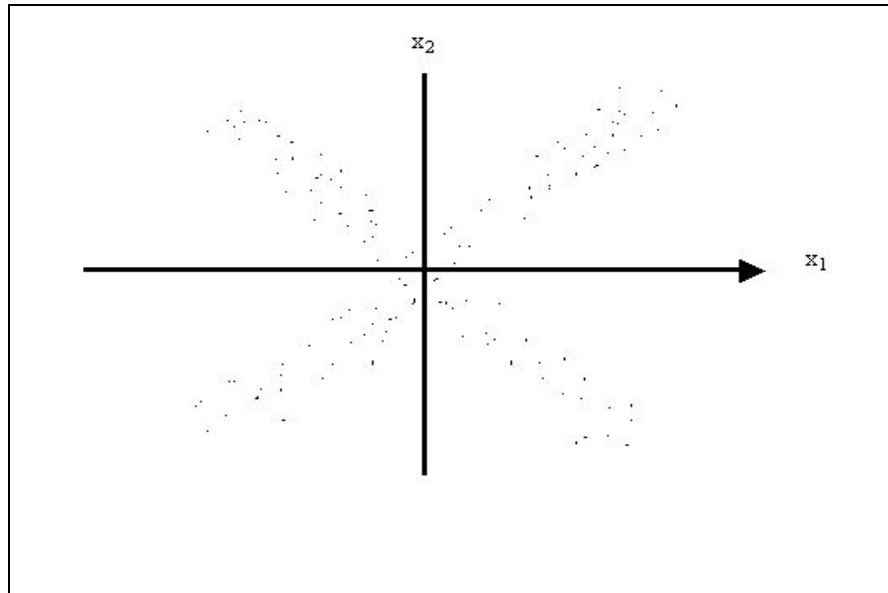
$$\mathbf{X}_{n \times m} = \mathbf{P}_a \mathbf{V}_a^T + \mathbf{P}_{m-a} \mathbf{V}_{m-a}^T \quad (2.10)$$

$$\mathbf{X}_{n \times m} = \mathbf{P}_a \mathbf{V}_a^T + \mathbf{E} \quad (2.11)$$

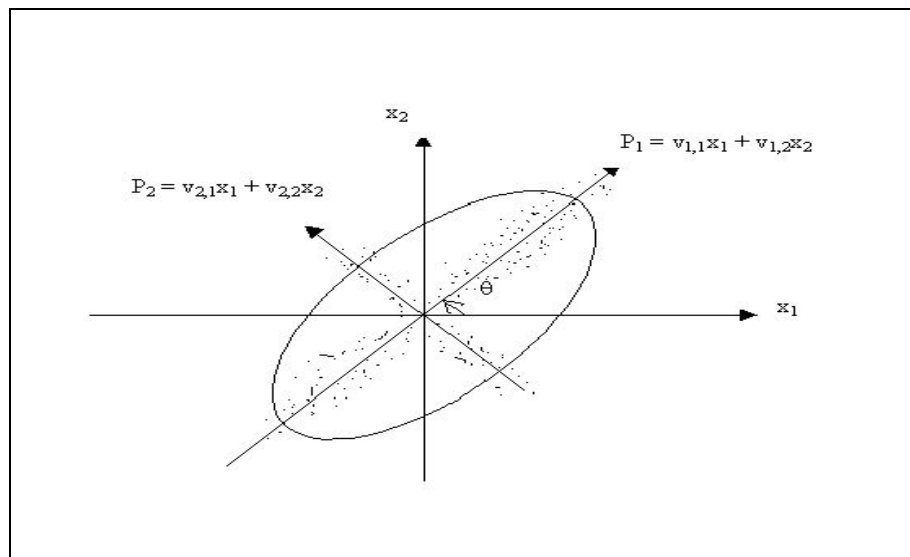
The retained principal components  $[p_1, p_2, \dots, p_a]$ , which form the  $\mathbf{P}_a \mathbf{V}_a^T$  term, are associated with systematic variation in data while the residual principal components  $[p_{a+1}, p_{a+2}, \dots, p_m]$ , which form the residual matrix  $\mathbf{E}$  are considered of containing measurement errors (Seborg *et al.*, 1996). Therefore, PCA is a multivariate analysis technique that could use less number of newly formed variables to represent the original data variations without losing significant information, in which, information here is referred to data variation.

For the graphical representation of PCA, the linear combinations of the original variables in forming the new variables are actually representing selection of a new coordinate system with  $[P_1, P_2, \dots, P_m]$  as the new axes obtained by rotating the original

system with  $x_1, x_2, \dots, x_n$  as the coordinate axes. The new axes represent the direction with maximum variability and provide a simpler and more parsimonious description of the variance-covariance matrix or correlation matrix (Johnson and Wichern, 1992). Figure 2.1 and 2.2 are prepared to give graphical representations of PCA



**Figure 2.1:** System for Two Variables Distribution



**Figure 2.2:** Graphical Representation of PCA

### 2.3 Fundamentals / Theory of Process Monitoring on MSPM Using PCA Tools

Reformation and upgrading of conventional Statistical Process Control (SPC) method has produce MSPC. MSPC tools such as principal component analysis (PCA) were used to reduce the explaining dimension of the process data. Maestri et al. say this method has show great success and particularly suited to data set comprising correlated and collinear variables. Ge and Song (2008) define process data as different group based, for instance, on variation in the operating capacity, seasonal variations or changes in the feedstock characteristics and also on modifications in the operation strategies. From a geometric point of view, whenever such as a change occurs, the process data tend to group into a new cluster in a different location in the high dimensional space containing the process normal operating region. However when the data is considered belong to a unique normal operating region, the volume of this region becomes incorrectly large. Zhao et al, (2006) say this region will lead to an increasing number of missing and false alarm. According to Zhao et al, (2004) when PCA based MSPC tools applied to a process with multiple operating modes, many missing and false alarm can appear even when the process itself is operating under other steady-state nominal operating conditions. Particularly this technique is for reducing the number of dimensions used from the original data as well as projected them into a number of uncorrelated variables, by means of forming the appropriate linear combinations of the original variables. Hence, MSPC is the only method where the data is treated simultaneously by way of reducing the dimensionality of the data observed without losing any of important information. In addition, this method can reduce the burden of constructing a large amount of single-variable control charts and enable detecting events that are impossible or difficult to detect from the single-variable control charts (Phatak, 1999).

According to Venkatasubramaniam et al, (2003) multivariate statistical techniques are powerful tool that capable to compressing data and reducing its dimensionality. Hence the essential information is retained and easy to analyze than the original huge data set. Moreover, it is able to handle noise and correlation to extract true information effectively. Initially, PCA method is proposed by Pearson (1901) later, it been develop by Hotelling (1947). This is a standard multivariate technique which has

been including in many textbooks (Jackson, 1991; Anderson, 1984) and research paper (Wold, Esbensen and Geladi, 1987; Wold, 1978). Venkatasubramaniam et al, (2003) say PCA is based on orthogonal decomposition of the covariance matrix of the process variables along directions that explain the maximum variation of the data. Yu and Zhang say this method involved a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables, which are called principal component.

## **2.4 Extension Of Principal Component Analysis**

There are many extension of Principle Component Analysis (PCA) which is some of these is Kernel of PCA, Multiway-PCA, Multi-Scale PCA, Moving PCA, Multi-Block PCA and many more.

### **2..4.1 Multiway-PCA**

Conventional PCA is best for analyzing a two-dimensional matrix of data collected from a steady state process, containing linear relationships between the variables. Since these conditions are often not satisfied in practice, several extensions of PCA have been developed. Nomikos and MacGregor (1994) proposed Multi-Way PCA, which allows the analysis of a multi-dimensional matrix. Multi way method organized the data into time ordered block which each represent a single sample or process run. Three dimensional array data (I, batch samples x J, process variables x K, time) is decomposed to two dimensional array (I x JK) data for easier analysis (Wise and Gallagher, 1996). Projection of these three dimensions data into two dimensions makes this method suitable and widely applied for batch processes. Nomikos and MacGregor (1994) used simulated data obtained from a semibatch reactor to monitor the process.

Multi-Way PCA was applied to industrial batch polymerization reactor using Hotelling's T<sup>2</sup> chart for fault detection and contribution plots for fault diagnosis